- Sanz, J. y Macías, A. (2004). "Dinámicas de calidad y cambio institucional en los sistemas de producción local de aceite de oliva", en *Congreso Internacional "Agroindustria Rural y Territorio".*- Toluca: Universidad Autónoma del Estado de México.

- Toledo, C. y Bartra, A. coords. (2000). *Del círculo vicioso al círculo virtuoso. Cinco miradas al desarrollo sustentable de las regiones marginadas.*–México, D.F.: Plaza y Valdés – SEMARNAP.

- Sylvander, Bertil (1996). " Normalisation et concurrence internationale: la politique de qualité alimentaire en Europe"; en *Economie Rurale*, no. 231, pp. 56–61.

- Valceschini, Egizio (1999). " Les signaux de qualité crédibles sur les marchés agroalimentaires: certifications officielles et marques "; en L. Lagrange (Coord.) *Signes officiels de qualité et développement agricole.*– Paris: Eds. Technique & Documentation, pp. 147–166.

The Duration of Civil Cases. A Survival Analysis

Carmen Vargas Pérez
Centro de Investigaciones en Derecho y Economía
Universidad Carlos III de Madrid
cvargas@eco.uc3m.es

Juan Luis Peñaloza Figueroa
Sección Departamental de Estadística e Investigación Operativa II
Universidad Complutense de Madrid
jluispf@emp.ucm.es

Abstract

Estimations about the duration of judicial cases are usually initiated by eliminating all those observations in which at least one of the two relevant dates (filing or termination of the procedure) are unknown. This method may lead to bias in estimations. Thus, in this paper we make an application of Survival Analysis to assess the duration of civil cases in Argentina (Federal Capital and Santa Fe). We use Kaplan-Meier Product Limit Method to estimate survival functions (and thus, duration probabilities) and then proceed to analyse the effect of several variables using a Cox Proportional Hazard Regression.

Key words: Duration of Case Processing, Survival Analysis, Kaplan Meier Estimates, Cox Regression Model
JEL Classification Codes: K00, K41, H14

Introduction

The importance given to the analysis of the duration of judicial procedures arises not only because of its economic and social consequences, but also because of its significance in terms of judicial policy. On the one hand, judicial delay, as a symptom of low performance, is considered as an increase of uncertainty and risk placed on investment, associated with higher costs and less profits of their activity. The first consequences will be price increases and supply reductions. In addition, employment, credits, building,.. will be reduced, and some "hidden costs" will be faced. [8] On the other hand, as measurement and diffusion of data about this problem is also scarce, public trust and confidence are affected. People usually think that "slow justice" and "unfair justice" are synonyms, and unfortunately, the less information citizens have, the worse is their opinion of the judicial system.

It is for this reason that precise estimates about the real duration of cases are so important, being urgent to know the main factors associated with that duration.

Given the frequent complexity of gathering this type of data, there are not so many studies of the duration of judicial cases filed, and among them, a usual feature is to constrain the sample to those cases where the actual time of case filing and case disposition is known. In the absence of additional procedures to make use of the partial information contained in the unfinished cases, it would be reasonable to discard the "censored" cases to guarantee data homogeneity. But, by removing them, information contained in them is eliminated, and so the fundamental statistical principle of using all the available information is not fulfilled.[9]

Research on epidemiology, biology or health related areas usually face problems regarding "lost" or incomplete observations. For instance, when studying survival time of individuals affected by some disease, as a function of medication taken, tobacco addiction or any sanitary measure applied, it is frequent that some individuals either abandon the study, or die (for reasons different to the ones under study) or, on the opposite

[8] About the different costs that low judicial performance produces see Pastor y Vargas (2001).

[9] Note that the longer is a case, the higher is its probability of be considered a censored one.

side, are still alive at the end of the period of data gathering. These cases (individuals) are called Censored. To avoid discarding the partial information contained in that data, Survival Analysis has usually been used, and thus considered to be a bio-statistical research tool.

In this paper we make an application of survival analysis techniques that let us make good use also of the partial information we have got about the claims that are brought to court; even if the case has not yet finished, we know that it will last at least the time it has been already in the tribunal.

In this paper, we apply survival analysis to the duration of civil cases filed in Argentina, in the Federal Capital and Santa Fe. Civil cases filed in 1996 were recorded, including Commercial, Labour, Administrative and Social Security claims. Sample size was 483 cases.[10]

The *Survival* of Judicial Cases

We first estimate Survival Functions[11] –in our case referring to duration– of the cases under study. We use the Kaplan-Meier Product Limit method. The sample was classified in five strata, corresponding to each of the five areas studied.

Case duration or survival time is measured in months, recording the time elapsed between case filing and case termination, or between filing and the last known date, for those lawsuits not ended by the time data was gathered –i.e. censored cases.

Table 1 shows the total number of observations recorded for each area, and the distribution of cases with complete duration data (Non censored) and those not yet finished (Censored).
The following table shows that median survival time ranges from 6.8 months in Administrative cases –with a 95% confidence interval between 8 and 11.7 months- to more than 14 months in labour claims, with an upper limit of 19 months.

The global survival function for the whole data set is shown below.

[10] The sample was obtained from a data base posted on "Databases on Uses and Users of Justice", World Bank (http://www1.worldbank.org/publicsector/legal/ empiricalresearch.htm). Original data was gathered by researchers at Fores (Argentina). A report containing the main results of that research work can be found in Garavano (2001).

[11] Survival Function: the probability that the survival time of an individual (case) is longer than some particular value.

Table 1: Censored and Non Censored Values

Area	N	Non Censored	Censored	% Censored
Civil	196	136	60	30.6
Commercial	46	41	5	10.9
Administrative	40	36	4	10.0
Labour	157	116	41	26.1
Soc. Security	44	42	2	4.5
Total	483	371	112	23.2

Table 2: Median Survival Times in the Five Areas

		Confidence Interval	
Area	Estimated Median	Lower L.	Upper L.
Civil	10.0	8.2	14.2
Commercial	11.0	7.3	15.0
Administrative	6.8	5.0	9.4
Labour	14.1	10.2	19.2
Soc. Security	9.8	8.1	11.7

In general, variation of survival with respect to time elapsed (since the claim was filed) is rather uniform, although it appears that the speed at which it decreases is slightly higher before the twentieth month.

Some specific information about survival by area, after 1.5 and 2 years, is shown in the table below.

Data shows that labour claims not only have the higher median among all the areas, but also that the probability of lasting more than 1,5 years is greater than 40%, and of lasting more than 2 years is near 30%. In fact, we have estimated that a duration probability greater than 30 months is also high (17%). The probability of longer cases is lower in the Civil area, but it is also important. As shown in the table, the probability of exceeding two years is 25%. On the opposite side, the survival function of cases brought to administrative courts fall quickly to 11% at the 18th

Figure 1: Kaplan-Meier Estimator. Survival of the Whole Set of Cases under Study.

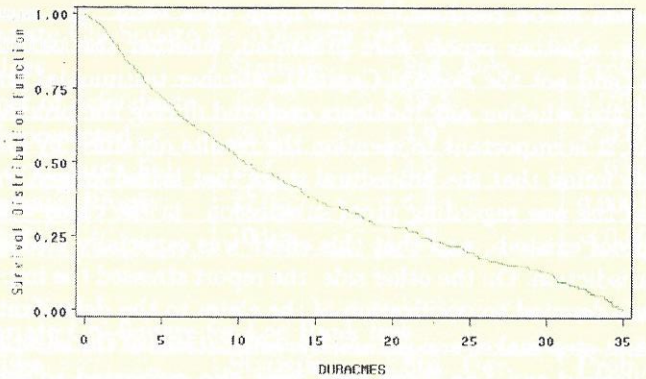

Table 3: Kaplan-Meier Estimators. Survival (probabilities) in the Five Areas.

	Survival Function Estimates			
Area	12 months	18 months	24 months	Std. Error
Civil	0.47	0.30	0.25	0.062
Commercial	0.43	0.25	0.06	0.062
Administrative	0.25	0.19	0.06	0.060
Labour	0.53	0.42	0.28	0.042
Soc. Security	0.32	0.11	0.04	0.051

month, and the probability of lasting longer is rather small (4%). It is important to stress that survival of cases filed in the Social Security tribunals quickly decrease –to 19%- after 18 months.[12]

[12]We also made comparisons among the different strata survival curves. Homogeneity tests –log rank test, Wilcoxon test and the Log Likelihood Ratio- show significant differences among the corresponding curves, all of them with significance levels greater than 90%. In particular, the Log Rank test –that shows a higher probability of detecting late differences- is more significant than the Wilcoxon text –that attaches great importance to short times- (p=0,0092 for the Log Rank test, p=0,0820 for the

Factors Significantly Related to Case Duration

We analysed a wide range of variables, potentially related with the time a case needs to be resolved.[13] The main ones were: the number of defendants, whether proofs were presented, whether the territory was Santa Fe (and not the Federal Capital), whether testimonial proof was presented and whether any incidents occurred during the processing. In particular, it is important to mention the results obtained by Garavano (2001). He found that the procedural stage that lasted longer, in all the areas, was the one regarding proof submission –in the cases where this type of proof existed-, and that this effect was especially strong in the labour jurisdiction. On the other side, the report stressed the importance of the time devoted to notification of the claim to the defendants.

The first step makes non parametric estimations of the survival functions. Wilcoxon and Log Rank tests were also estimated to assess the association of case survival with the several explaining variables considered. Of course, stratification by area (Civil, Commercial, Administrative ...) was preserved in this step, so the estimators control for this variable.

The results of univariate Wilcoxon and Log Rank tests that marginally assess the effect of every factor are shown below. It can be verified that every variable is individually significant in both tests, with the only exception of the "number of defendants" according with the second test (Pr= 0.1873).

After that, we start by including the one with the highest Chi-Square and continue adding the rest, in a sequential way, by using the highest marginal contribution criterion. The results of this procedure, considering both the Wilcoxon and the Log Rank tests, are shown below.

As shown, global Chi Square statistic reaches values of 128.6 and 99.2 for the Wilcoxon and Log Rank tests respectively; that is, both are highly significant (p<0,0001).

The second step applies a Cox Proportional Hazard Regression to study the effect of explanatory variables. As we mentioned before, we had

Wilcoxon test and p=0,0126 for the Log Likelihood Ratio).

[13]Only those variables that showed the greatest association are mentioned in the text, discarding also those that resulted to be almost exact linear combinations of the rest. The full list of variables considered is shown in the Annex.

Table 4: Rank Tests of Association of Survival Functions with Explaining Variables.

Univariate Chi-Square for Wilcoxon test			
Variables	Statistic	Std. Dev.	Prob. Value
Number of defendants	43.4	19.4	0.0250
Proofs presented	39.0	4.9	< 0.0001
Territory (Santa Fe)	28.1	4.9	< 0.0002
Testimonial Proofs	27.0	4.3	0.0001
Incidents	10.9	2.8	< 0.0001

Univariate Chi-Square for Log Rank test			
Variables	Statistic	Std. Dev.	Prob. Value
Number of defendants	56.8	43.0	0.1873
Proofs presented	58.7	8.5	< 0.0001
Territory (Santa Fe)	42.5	7.1	< 0.0001
Testimonial Proofs	37.5	7.3	< 0.0001
Incidents	15.4	5.4	0.0045

complete information about duration for 371 of the 483 cases considered. For the rest, 112, only partial information was recorded, as their state was "unfinished" and we only know how long the proceeding had been in the court when the data was gathered. The variables considered to be associated with survival times are the same considered in the previous sections.[14]

The variables were included by a forward selection method, estimating the adjusted chi square statistic for each of the variables that had not yet been included, considering a minimum significance level of 0.10 for any of them to be included in the model. In addition, Hazard Ratios and confidence intervals were estimated for every variable.[15]

Proofs presented was the first variable introduced, having a Chi

[14]To simplify the presentation, the details about the analysis of residuals made to detect a possible lack of fitness of the model are omitted.

[15]Score test is used.

Table 5: Forward Chi Square Sequence for Wilcoxon and Log Rank Tests.

Chi Square Sequence for Wilcoxon Test				
Variables	Partial Chi-Square	Probability Value	Chi Increment	Increment Probability
Proofs presented	64.3	< 0.0001	64.3	< 0.0001
Territory (Santa Fe)	108.0	< 0.0001	43.6	< 0.0001
Incidents	116.9	< 0.0001	8.86	0.0029
Number of defendants	123.8	< 0.0001	6.96	0.0083
Testimonial Proofs	128.6	< 0.0001	4.82	0.0281

Chi Square Sequence for Log Rank Test				
Variables	Partial Chi-Square	Probability Value	Chi Increment	Increment Probability
Proofs presented	47.7	< 0.0001	47.6	< 0.0001
Territory (Santa Fe)	87.3	< 0.0001	39.6	< 0.0001
Incidents	93.3	< 0.0001	6.0	0.0143
Number of defendants	96.4	< 0.0001	3.1	0.0771
Testimonial Proofs	99.2	< 0.0001	2.7	0.0955

Square statistic of 47.67 and a Pr-value lower than 0.0001. The three tests applied were significant, thus rejecting the null hypothesis (betas=0).[16] The next step led to the introduction of the Territory variable. The Number of Defendants was the last variable introduced in the model.

The final results of the Log likelihood estimations, after introducing the above mentioned variables, are shown in Table 6.[17]

The values in Table 6 show that hazard ratio for every variable is less than one, and the estimated parameters are all negative, thus implying that all of them are *positively* associated with survival in courts. That is, the three variables (proof presentation, claim filing in Santa Fe and

[16] Every step estimated the following global tests: Likelihood Ratio, Score test and Wald test. Significance of coefficients was also checked.

[17] Final values of global significance tests for the model were 102.7, 96.43 and 90.54 for the Likelihood Ratio, Score test and Wale test, respectively, all of them with Pr<0,0001.

Table 6: Final Results of Log Likelihood Estimations.

	Estimated Parameter	Std. Error	Wald Chi Square
Proofs presented	-1.007	0.139	52.69
Territory (Santa Fe)	-0.872	0.136	40.90
Incidents	-0.492	0.247	3.98
Number of defendants	-0.084	0.052	2.59

	Prob. Value	Hazard Ratio	Confidence Interval	
			Lower Limit	Upper Limit
Proofs presented	< 0.0001	0.365	0.278	0.480
Territory (Santa Fe)	< 0.0001	0.418	0.320	0.546
Incidents	0.046	0.611	0.377	0.991
Number of defendants	0.107	0.920	0.831	1.018

the occurrence of incidents) have a positive and significant effect on the duration of the civil lawsuits studied.

Conclusions

The purpose of this paper was to study the duration of civil cases brought to Argentinean courts in the Federal Capital and Santa Fe, and to assess the effect of several variables. As the sample contained observations with complete information about the duration of the cases, but also censored data, we chose Survival Analysis and Cox Proportional Hazard Regression as our research tools.

Although Survival Analysis has mainly been considered a bio statistical tool in the past, we think it is an essential procedure in studies like this one, in which part of the data is always censored. In fact, precisely because of the technical difficulties involved in this type of data, a usual sample design method is to constrain the analysis to the sub sample of cases already finished at the time of data gathering. Thus, every case whose date of judgement is unknown or not yet terminated is discarded.

In this paper we have applied techniques for censored data to avoid

the usual bias generated when eliminating all the cases where filing or termination dates are not known. Area stratification showed, with high significance levels, the difference among the corresponding survival functions. Survival analysis also showed the main variables that could possibly explain survival changes.

In a second phase of analysis a Cox Proportional Hazard Model was applied. After checking its suitableness, we proceeded to assess the effect of several variables that could possibly be associated with survival of cases in the court. The four variables that showed a significant effect (Number of Defendants, Proof Presentation, Incidents and Santa Fe Territory) were positively associated with claim duration.

Part of these results coincide with those by Garavano (2001) who suggested that defendants may tend to increase case duration because of strategic reasons, finding that Notification of Claims (a task that can be delayed by the defendant's attorney) is one of the longest stages of civil procedures. In our study that hypothesis gains reliability, as we find that the Number of Defendants has a positive effect. This author also associates judicial delay with the submission of the proofs, confirming the great influence that plaintiffs may have on duration, and thus suggests that judges should put more control of the procedure (and its duration) in their hands. In addition to these variables, our analysis suggests that cases coming from Santa Fe are longer and that the same positive effect on duration comes from the information of incidents. We think that additional local studies should be made to determine the source of that delay and to assess the possibility of reducing it.

We hope this paper contributes to the study of courts performance and the methods to estimate the duration of claims that come before the courts. We are confident that the techniques here applied will be an incentive to policy makers to create judicial data bases with systematic and reliable data on the total time needed for a claim heard in court to get a final judgement. And that this study will encourage other researchers to enlarge their samples sizes by considering the information included in censored data. That will lead to less biased estimations about duration of procedures and a more reliable identification of factors associated with it.

ANNEX. Explanatory variables considered at the beginning of the study

- Area: (Civil, Commercial, Labour, Administrative, Social Security)
- Territory (Santa Fe or Federal Capital)
- Territorial Jurisdiction (Federal, of Federal Capital; National of Federal Capital; Province of Santa Fe)
- Number of participants
- Procedural Method (Oral or Written)
- Type of Procedure
- Money claim?
- Number of judicial procedures
- Number of non suits
- Testimonial Proofs?
- Informative Proofs?
- Number of appeals
- Result of each appeal
- Additional measures?
- Incidents?
- Number of Incidents
- Related cases?

References

1. Fenn, P. and N. Rickman (1999). "Delay and Settlement in Litigation", *Economic Journal*, July, Vol. 109, iss. 457, pp. 476-91.

2. Garavano, G. (2001). *Los Usuarios del Sistema de Justicia en Argentina.* Argentina. Fores.

3. Helland, E. and A. Tabarrok (2003). "Contingency Fees, Settlement Delay, and Low-Quality Litigation: Empirical Evidence from Two Datasets", *Journal of Law, Economics, and Organization*, Special Issue, October, Vol. 19, iss. 2, pp. 517-42.

4. Hughes, J. W. and E. Savoca (1999). "Accounting for censoring in duration data. An application to estimating the effect of legal reforms on the duration of medical malpractice disputes". *Journal of Applied Statistics*, Vol. 26, iss. 2, pp. 219-228.

5. McWilliams, J. M. (1992). "Setting the Record Straight: Facts about Litigation Costs and Delay", *Business Economics*, October, Vol. 27, iss. 4, pp. 19-24.

6. Pastor, S., C. Vargas, J. Hasbún and G. de Jesús (2000). *Informe Estadístico y Cualitativo sobre los Usuarios de la Justicia en la República Dominicana en las Áreas Civil Comercial y Laboral.* República Dominicana. Banco Mundial and Fundación Institucionalidad y Justicia.

7. Pastor, S. and C. Vargas (2001). "El coste de la Justicia. Datos y un poco de análisis", *Cuadernos de Derecho Judicial.* Vol. XV, pp. 29-72, Madrid.

8. Pastor, S (2003). "Eficiencia y eficacia de la justicia". (with English summary). *Papeles de Economía Española*, iss. 95, pp. 272-305.

9. Torre, A (2003). "The Impact of Court Delays on the Prosecutor and the Defendant: An Economic Analysis"; *European Journal of Law and Economics*, July, Vol. 16, iss. 1, pp. 91-111.

10. Vargas, C. and J. L. Peñaloza (2004). "Qué Costes Económicos y Sociales Comporta la Ejecución de Sentencias Judiciales". *Cuadernos de Estudios Empresariales*, Vol. 14.

11. Vereeck, L and M. Muhl. "An Economic Theory of Court Delay" *European Journal of Law and Economics,* November 2000, Vol. 10, iss. 3, pp. 243-68, Madrid.